How to determine if a random graph with a fixed degree sequence has a giant component

Felix Joos* Guillem Perarnau† Dieter Rautenbach‡ Bruce Reed§

June 30, 2016

Abstract

The traditional Erdős-Rényi model of a random network is of little use in modelling the type of complex networks which modern researchers study. In this graph, every pair of vertices is equally likely to be connected by an edge. However, 21st century networks are of diverse nature and usually exhibit inhomogeneity among their nodes and correlations among their edges. This motivates the study, for a fixed degree sequence $\mathcal{D} = (d_1, ..., d_n)$, of a uniformly chosen simple graph $G(\mathcal{D})$ on $\{1, ..., n\}$ where the vertex i has degree d_i. In this paper, we study the existence of a giant component in $G(\mathcal{D})$.

A heuristic argument suggests that a giant component in $G(\mathcal{D})$ will exist provided that the sum of the squares of the degrees is larger than twice the sum of the degrees. In 1995, Molloy and Reed essentially proved this to be the case when the degree sequence $\mathcal{D}$ under consideration satisfies certain technical conditions [*Random Structures & Algorithms*, **6**:161–180]. This work has attracted considerable attention, has been extended to degree sequences under weaker conditions and has been applied to random models of a wide range of complex networks such as the World Wide Web or biological systems operating at a sub-molecular level. Nevertheless, the technical conditions on $\mathcal{D}$ restrict the applicability of the result to sequences where the vertices of high degree play no important role. This is a major problem since it is observed in many real-world networks, such as scale-free networks, that vertices of high degree (the so-called *hubs*) are present and play a crucial role.

In this paper we characterize when a uniformly random graph with a fixed degree sequence has a giant component. Our main result holds for every degree sequences of length n such that the sum of the degrees in the sequence which are not 2 is at least $\lambda(n)$ for some arbitrary function λ going to infinity with n. Besides the fact that it is a minor technical condition, the typical structure of $G(\mathcal{D})$, when $\mathcal{D}$ does not satisfy it, is relatively simple and easy to understand.

Our result gives a unified criterion that implies all the known results on the existence of a giant component in $G(\mathcal{D})$, including both the generalizations of the Molloy-Reed result and the results on more restrictive models such as the Aiello-Chung-Lu Power-Law random graph [*STOC 2000*, 171–180]. Moreover, it turns out that the heuristic argument used in all the previous works on the topic, does not extend to ~~actually suggests the wrong answer for~~ general degree sequences.

*School of Mathematics, University of Birmingham, Birmingham, B15 2TT, UK. E-mail: f.joos@bham.ac.uk. This coauthor was supported by the EPSRC, grant no. EP/M009408/1.

†School of Mathematics, University of Birmingham, Birmingham, B15 2TT, UK. E-mail: p.melliug@gmail.com. This coauthor was supported by the European Research Council, ERC Grant Agreement no. 306349.

‡Institute of Optimization and Operations Research, Ulm University, Ulm, Germany. E-mail: dieter.rautenbach@uni-ulm.de

§Laboratoire I3S, CNRS, Sophia Antipolis, France and Visiting Research Professor, Kawarabayashi Large Graph ERATO Project, NII, Japan; E-mail: reed@i3s.unice.fr

1 Introduction

The traditional Erdős-Rényi model of a random network is of little use in modelling the type of complex networks which modern researchers study. Such a graph can be constructed by adding edges one by one such that in every step, every pair of non-adjacent vertices is equally likely to be connected by the new edge. However, 21st century networks are of diverse nature and usually exhibit inhomogeneity among their nodes and correlations among their edges. For example, we observe empirically in the web that certain authoritative pages will have many more links entering them than typical ones. This motivates the study, for a fixed degree sequence $\mathcal{D} = (d_1, ..., d_n)$, of a uniformly chosen simple graph $G(\mathcal{D})$ on $[n] = \{1, ..., n\}$ where the vertex i has degree d_i. In this paper, we study the existence of a giant component in $G(\mathcal{D})$.

A heuristic argument suggests that a giant component will exist provided that the sum of the squares of the degrees is larger than twice the sum of the degrees. In 1995, Molloy and Reed essentially proved this to be the case provided that the degree sequence under consideration satisfied certain technical conditions [24]. This work has attracted considerable attention and has been applied to random models of a wide range of complex networks such as the World Wide Web or biological systems operating at a sub-molecular level [1, 2, 5, 28, 29]. Furthermore, many authors have obtained related results which formalize the Molloy-Reed heuristic argument under different sets of technical conditions [6, 15, 18, 21, 25].

Unfortunately, these technical conditions do not allow the application of such results to many degree sequences that describe real-world networks. While these conditions are of different nature, here we exemplify their limitations with a well-known example, scale-free networks. A network is *scale-free* if its degree distribution follows a power-law, governed by an exponent. It is well-known that many real-world networks are scale-free and one of the main research topic in this area is to determine the exponent of a particular network. It has been observed that many scale-free networks have a fat-tailed power-law degree distribution with exponent between 2 and 3. This is the case of the World Wide Web, where the exponent is between 2.15 and 2.2 [9], or the Movie Actor network, with exponent 2.3 [4]. In scale-free networks with exponents between 2 and 3, the vertices of high degree (called *hubs*) have a crucial role in several of the network properties, such as in the "small-world" phenomenon. However, one of the many technical conditions under the previous results on the existence of a giant component in $G(\mathcal{D})$ hold, is that the vertices of high degree do not have a large impact on the structure of the graph. (In particular, it is required that there is no mass of edges in vertices of non-constant degree.) Hence, often these results cannot be directly applied to real-world networks where hubs are present and for each particular network ad-hoc approaches are needed (see e.g. the Aiello-Chung-Lu model for the case of scale-free networks [1]).

Another problem is that all the previous results apply to a sequence of degree sequences $(\mathcal{D}_n)_{n\geq 1}$ instead of a degree sequence $\mathcal{D}$ of fixed length. This can be a major problem when modelling complex networks. In most of the real-world applications, researchers extract the degree sequence $\mathcal{D}$ of a particular network and then aim to model the structure of such network by considering a random network with degree sequence $\mathcal{D}$. Since the previous results are on sequences of degree sequences, it is possible that they give no information for the particular degree sequence $\mathcal{D}$.

Finally, all the previous results on the existence of a giant component in $G(\mathcal{D})$ do not cover degree sequences where most of the vertices have degree 2.

In this paper we characterize when $G(\mathcal{D})$ has a giant component for *every* degree sequence $\mathcal{D}$ of length n. We only require that the sum of the degrees in the sequence which are not 2 is at least $\lambda(n)$ for some arbitrary function λ going to infinity with n. Besides the fact that it is a relatively minor technical condition, we also show that if it is not satisfied, both the probability that $G(\mathcal{D})$ has a giant component and the probability that $G(\mathcal{D})$ has no giant component are bounded away

from 0.

It turns out that the heuristic argument which was used in [24] to describe the existence of a giant component in $G(\mathcal{D})$ for degree sequences satisfying some technical conditions and that was generalized in the subsequent papers [6, 15, 18, 21, 25], does not extend to ~~actually suggests the wrong answer for~~ general degree sequences. Precisely, if we let S be a smallest set such that (i) no vertex outside of S has degree bigger than a vertex in S, and (ii) the sum of the squares of the degrees of the vertices outside of S is at most twice the sum of their degrees, then whether or not a giant component exists depends on the sum of the degrees of the vertices in S, not on the sum of the squares of the degrees of the vertices in S as suggested by this heuristic argument

This new unified criterion on the existence of a giant component in $G(\mathcal{D})$ is valid for every sequence $\mathcal{D}$ and implies all the previous results on the topic both for arbitrary degree sequences [6, 18, 24] or for particular models [1].

In this paper we present our main results and briefly describe the proof strategy. We refer the interested reader to the full version of the paper [19] for complete proofs.

1.1 The Molloy-Reed Approach

Let us first describe the result of Molloy and Reed [24]. Throughout the paper we assume that all the d_i are positive, as we can simply delete the isolated vertices from the graph and analyse what remains. It is straightforward to transfer our results to the case when there are vertices of degree 0. We also restrict our attention to feasible degree sequences, that is, those $\mathcal{D}$ such that the set of graphs with degree sequence $\mathcal{D}$ is nonempty.

For every $1 \leq i \leq n$, one can explore the component containing a specific initial vertex i of a graph on $[n]$ via breadth-first search. Initially we have d_i "open" edges out of i. Upon exposing the other endpoint j of such an open edge, it is no longer open, but we gain $d_j - 1$ open edges out of j. Thus, the number of open edges has increased by $d_j - 2$ (note that this is negative if $d_j = 1$).

One can generate the random graph $G(\mathcal{D})$ for $\mathcal{D} = (d_1, ..., d_n)$ and carry out this exploration at the same time, by choosing each vertex as j with the appropriate probability.

Intuitively speaking, the probability we pick a specific vertex j as the other endpoint of the first exposed edge is proportional to its degree. So, the expected increase in the number of open edges in the first step is equal to $\frac{\sum_{k\in[n]\setminus\{i\}} d_k(d_k-2)}{\sum_{k\in[n]\setminus\{i\}} d_k}$. Thus, it is positive essentially if and only if the sum of the squares of the degrees exceeds twice the sum of the degrees.

Suppose that this expected increase remains the same until we have exposed a linear number of vertices. It seems intuitively clear that if the expected increase is less than 0, then the probability that initial vertex i is in a linear order component is very small, and hence the probability that $G(\mathcal{D})$ has no linear order component is $1 - o(1)$. If for some positive constant ϵ, the expected increase is at least ϵ, then there is some $\gamma = \gamma(\epsilon) > 0$ such that the probability that i is in a component with at least γn vertices exceeds γ.

In [24], Molloy and Reed proved, subject to certain technical conditions which required them to discuss sequences of degree sequences rather than one single degree sequence, that we essentially have that (i) if $\frac{\sum_{k=1}^{n} d_k(d_k-2)}{\sum_{k=1}^{n} d_k} > \epsilon$ for some $\epsilon > 0$, then the probability that $G(\mathcal{D})$ has a giant component is $1-o(1)$, and (ii) if $\frac{\sum_{k=1}^{n} d_k(d_k-2)}{\sum_{k=1}^{n} d_k} < -\epsilon$ for some $\epsilon > 0$, then the probability that $G(\mathcal{D})$ has no giant component is $1 - o(1)$. We present their precise result and some of its generalizations later in this introductory section.

1.2 Our Refinement

It turns out that, absent the imposed technical conditions, the expected increase may change drastically during the exploration process. Consider for example the situation in which $n = k^2$ for some large odd k, $d_1 = d_2 = .. = d_{n-1} = 1$ and $d_n = 2k$. Then $\frac{\sum_{k=1}^{n} d_k(d_k-2)}{\sum_{k=1}^{n} d_k} = \frac{4k^2-4k-(n-1)}{2k+n-1} \approx 3$, and so the Molloy-Reed approach would suggest that with probability $1 - o(1)$ there will be a giant component. However, with probability 1, $G(\mathcal{D})$ is the disjoint union of a star with $2k$ leaves and $\frac{n-2k-1}{2}$ components of order 2 and hence it has no giant component. The problem is that as soon as we explore vertex n, the expected increase drops from roughly 3 to -1, so it does not stay positive throughout the process.

Thus, we see that the Molloy-Reed criterion cannot be extended for general degree sequences. To find a variant which applies to arbitrary degree sequences, we need to characterize those for which the expected increase remains positive for a sufficiently long time.

Intuitively, since the probability that we explore a vertex is essentially proportional to its degree, in lower bounding the length of the period during which the expected increase remains positive, we could assume that the exploration process picks at each step a highest degree vertex that has not been explored yet. Moreover, note that vertices of degree 2 have a neutral role in the exploration process as exposing such a vertex does not change the number of open edges. These observations suggest that we should focus on the following invariants of $\mathcal{D}$ defined by considering a permutation π of the vertices that satisfies $d_{\pi_1} \leq ... \leq d_{\pi_n}$:

- $j_{\mathcal{D}} = \min\left(\left\{j : j \in [n] \text{ and } \sum_{i=1}^{j} d_{\pi_i}(d_{\pi_i} - 2) > 0\right\} \cup \left\{n\right\}\right)$,
- $R_{\mathcal{D}} = \sum_{i=j_{\mathcal{D}}}^{n} d_{\pi_i}$, and
- $M_{\mathcal{D}} = \sum_{d_i \neq 2} d_i$.

We emphasize that these invariants are determined by the multiset of the degrees given by $\mathcal{D}$ and are independent from π.

Our intuition further suggests that in the exploration process, the expected increase in the number of open edges will be positive until we have explored $R_{\mathcal{D}}$ edges and will then become negative. Thus, we might expect to explore a component with about $R_{\mathcal{D}}$ edges, and indeed we can show this is the case.

This allows us to prove our main result which is that whether $G(\mathcal{D})$ has a giant component essentially depends on whether $R_{\mathcal{D}}$ is of the same order as $M_{\mathcal{D}}$ or not. There is however a caveat, this is not true if essentially all vertices have degree 2.

For any function $\lambda : \mathbb{N} \to \mathbb{N}$, we say a degree sequence $\mathcal{D}$ is *λ-well-behaved* or simply *well-behaved* if $M_{\mathcal{D}}$ is at least $\lambda(n)$. Our main results hold for any function $\lambda \to \infty$ as $n \to \infty$.

Theorem 1. *For any function $\delta \to 0$ as $n \to \infty$, for every $\gamma > 0$, if $\mathcal{D}$ is a well-behaved degree sequence with $R_{\mathcal{D}} \leq \delta(n) M_{\mathcal{D}}$, then the probability that $G(\mathcal{D})$ has a component of order at least γn is $o(1)$.*

Theorem 2. *For any positive constant ϵ, there is a $\gamma > 0$, such that if $\mathcal{D}$ is a well-behaved degree sequence with $R_{\mathcal{D}} \geq \epsilon M_{\mathcal{D}}$, then the probability that $G(\mathcal{D})$ has a component of order at least γn is $1 - o(1)$.*

As we shall see momentarily, previous results in this field apply to sequences of degree sequences, and required that these degree sequences approached a limit in some smooth way. We can easily deduce results for every sequence of degree sequences from the two theorems above, and from our results on degree sequences which are not well-behaved, presented in the next section.

We denote by $\mathfrak{D} = (\mathcal{D}_n)_{n\geq 1}$ a sequence of degree sequences where $\mathcal{D}_n$ has length n. We say that $\mathfrak{D}$ is *well-behaved* if for every b, there is an n_b such that for all $n > n_b$, we have $M_{\mathcal{D}_n} > b$; $\mathfrak{D}$ is *lower bounded* if for some $\epsilon > 0$, there is an n_ϵ such that for all $n > n_\epsilon$, we have $R_{\mathcal{D}_n} \geq \epsilon M_{\mathcal{D}_n}$; and $\mathfrak{D}$ is *upper bounded* if for every $\epsilon > 0$, there is an n_ϵ such that for all $n > n_\epsilon$, we have $R_{\mathcal{D}_n} \leq \epsilon M_{\mathcal{D}_n}$.

The following is an immediate consequence of Theorem 1 and 2, and Theorem 6, which will be presented in the next section.

Theorem 3. *For any well-behaved lower bounded sequence of degree sequences $\mathfrak{D}$, there is a $\gamma > 0$ such that the probability that $G(\mathcal{D}_n)$ has a component of order at least γn is $1 - o(1)$.*

For any well-behaved upper bounded sequence of degree sequences $\mathfrak{D}$ and every $\gamma > 0$, the probability that $G(\mathcal{D}_n)$ has a component of order at least γn is $o(1)$.

If a sequence of degree sequences $\mathfrak{D}$ is either not well behaved or neither upper bounded nor lower bounded, then for every sufficiently small positive γ, there is a $0 < \delta < 1$ such that there are both arbitrarily large n for which the probability that $G(\mathcal{D}_n)$ has a component of order at least γn is at least δ, and arbitrarily large n for which the probability that $G(\mathcal{D}_n)$ has a component of order at least γn is at most $1 - \delta$.

1.3 The Special Role of Vertices of Degree 2

At first glance, it may be surprising that the existence of a giant component depends on the ratio between $R_{\mathcal{D}}$ and $M_{\mathcal{D}}$ rather than the ratio between $R_{\mathcal{D}}$ and $\sum_{i=1}^{n} d_i$. It may also be unclear why we have to treat differently degree sequences where the sum of the degrees which are not 2 is bounded.

To clarify why our results are stated as they are, we now highlight the special role of vertices of degree 2.

We let $H(\mathcal{D})$ be the multigraph obtained from $G(\mathcal{D})$ by deleting all cyclic components[1] and suppressing all vertices of degree 2.[2] Clearly, $H(\mathcal{D})$ is uniquely determined by $G(\mathcal{D})$. Moreover, the degree sequence of $H(\mathcal{D})$ is precisely that of $G(\mathcal{D})$ without the vertices of degree 2. Note that the non-cyclic components of $G(\mathcal{D})$ can be obtained from the multigraph $H(\mathcal{D})$ by subdividing some of its edges, so that every loop is subdivided at least twice, and all but at most one edge of every set of parallel edges is subdivided at least once.

The number of vertices of a non-cyclic component of $G(\mathcal{D})$ equals the sum of the number of vertices of the corresponding component of $H(\mathcal{D})$ and the number of vertices of degree 2 used in subdividing its edges. Intuitively, the second term in this sum depends on the proportion of the edges in the corresponding component of $H(\mathcal{D})$. Subject to the caveat mentioned above and discussed below, if the number of vertices of degree 2 in $G(\mathcal{D})$ is much larger than the size[3] of $H(\mathcal{D})$, then the probability that $G(\mathcal{D})$ has a giant component is essentially the same as the probability $H(\mathcal{D})$ has a component containing a positive fraction of its edges. The same is true, although not as immediately obvious, even if the number of vertices of degree 2 is not this large.

[1] A component is *cyclic* if it is a cycle and *non-cyclic* if it is not.

[2] Here and throughout the paper, when we say we suppress a vertex u of degree 2, this means we delete u and we add an edge between its neighbours. Observe that this may create loops and multiple edges, so the resulting object might not be a simple graph.

[3] As it is standard, we use *order* and *size* to denote the number of vertices and the number of edges of a graph, respectively.

Theorem 4. *For every $\gamma > 0$, there exists a $\rho > 0$ such that for every well-behaved degree sequence $\mathcal{D}$, the probability that $G(\mathcal{D})$ has a component of order at least γn and $H(\mathcal{D})$ has no component of size at least $\rho M_{\mathcal{D}}$ is $o(1)$.*

Theorem 5. *For every $\rho > 0$, there exists a $\gamma > 0$ such that for every well-behaved degree sequence $\mathcal{D}$, the probability that $G(\mathcal{D})$ has no component of order at least γn and $H(\mathcal{D})$ has a component of size at least $\rho M_{\mathcal{D}}$ is $o(1)$.*

As we mentioned above, if $\mathcal{D}$ is not well-behaved, then the results conclusions in Theorem 1 and 2 do no longer hold. For instance, suppose that $M_{\mathcal{D}} = 0$, that is, $d_i = 2$ for every $i \in [n]$. Then $H(\mathcal{D})$ is empty and $G(\mathcal{D})$ is a uniformly chosen disjoint union of cycles. In this case it is known that the probability of having a giant component is bounded away both from 0 and 1 (see e.g. [3]). Indeed, the latter statement also holds whenever $M_{\mathcal{D}}$ is at most a constant.

Theorem 6. *For every $b \geq 0$ and every $0 < \gamma < 1$, there exists an $n_{b,\gamma}$ and a $0 < \delta < 1$ such that if $n > n_{b,\gamma}$ and $\mathcal{D}$ is a degree sequence with $M_{\mathcal{D}} \leq b$, then the probability that there is a component of order at least γn in $G(\mathcal{D})$ lies between δ and $1 - \delta$.*

This theorem both explains why we concentrate on well-behaved degree sequences and sets out how degree sequences which are not well-behaved actually behave (badly obviously). Combining it with Theorem 1 and 2 immediately implies Theorem 3. ~~We omit the straightforward details.~~

1.4 Previous Results

The study of the existence of a giant component in random graphs with an arbitrary prescribed degree sequence[4], started with the result of Molloy and Reed [24]. Although they define the concept of asymptotic degree sequences, we will state all the previous results in terms of sequences of degree sequences $\mathfrak{D} = (\mathcal{D}_n)_{n\geq 1}$. Using a symmetry argument, one can translate results for sequences of degree sequences to asymptotic degree sequences, and vice versa. For every $\mathcal{D}_n = (d_1, \ldots, d_n)$, we define $n_i = n_i(n) = |\{j : d_j = i\}|$.

Before stating their result, we need to introduce a number of properties of sequences of degree sequences. A sequence of degree sequences $\mathfrak{D}$ is

- *feasible*, if for every $n \geq 1$, there exists at least one simple graph on n vertices with degree sequence $\mathcal{D}_n$.
- *smooth*, if for every $i \geq 0$, there exists λ_i such that $\lim_{n\to\infty} \frac{n_i}{n} = \lambda_i$.
- *sparse*, if there exists $0 < \lambda < \infty$ such that $\lim_{n\to\infty} \sum_{i\geq 1} \frac{in_i}{n} = \lambda$.
- *f-bounded*, for some function f of n, if $n_i = 0$ for every $i > f(n)$.

In particular, observe that random graphs $G(\mathcal{D}_n)$ arising from a sparse sequence of degree sequences $\mathfrak{D}$ have a linear number of edges, provided that n is large enough.
Given that $\mathfrak{D}$ is smooth, we define the following parameter,

$$Q(\mathfrak{D}) = \sum_{i\geq 1} i(i-2)\lambda_i \ .$$

Note that $Q(\mathcal{D})$ is very close to the notion of initial expected increase described in Section 1.1.

A sequence of degree sequences $\mathfrak{D}$ satisfies the **MR**-*conditions* if

[4]Random graphs with special degree sequences had been studied earlier (see, e.g. [22, 34]).

(a.1) it is feasible, smooth and sparse,

(a.2) it is $n^{1/4-\epsilon}$-bounded, for some $\epsilon > 0$,

(a.3) for every $i \geq 1$, $\frac{i(i-2)n_i}{n}$ converges uniformly to $i(i-2)\lambda_i$, and

(a.4) $\lim_{n\to\infty} \sum_{i\geq 1} i(i-2)\frac{n_i}{n}$ exists and converges uniformly to $\sum_{i\geq 1} i(i-2)\lambda_i$.

For a precise statement of the uniform convergence on conditions (a.3)–(a.4), we refer the reader to [24].

Now we can precisely state the result of Molloy and Reed [24].

Theorem 7 (Molloy and Reed [24])**.** *Let $\mathfrak{D} = (\mathcal{D}_n)_{n\geq 1}$ be a sequence of degree sequences that satisfies the* **MR***-conditions. Let G be a graph with n vertices chosen uniformly at random among all such graphs with degree sequence $\mathcal{D}_n$. Then,*

1. *if $Q(\mathfrak{D}) > 0$, then there exists a constant c_1 such that the probability that G has a component of order at least $c_1 n$ is $1 - o(1)$.*
2. *if $Q(\mathfrak{D}) < 0$ and the sequence is $n^{1/8-\epsilon}$-bounded for some $\epsilon > 0$, then for every constant c_2, the probability that G has no component of order at least $c_2 n$ is $1 - o(1)$.*

Theorem 7 has been generalized to other sequences of degree sequences, which in particular include the case $Q(\mathfrak{D}) = 0$. Theorem 3 implies all the criteria for the existence of a giant component in $G(\mathcal{D}_n)$ introduced below (see [19]). However, while some of these results give a more precise description on the order of the largest component, our results only deal with the existential question.

A sequence of degree sequences $\mathfrak{D}$ satisfies the **JL**-*conditions* if

(b.1) is feasible, smooth, and sparse,

(b.2) $\sum_{i\geq 1} i^2 n_i = O(n)$, and

(b.3) $\lambda_1 > 0$.

Observe that if $\mathfrak{D}$ satisfies the **JL**-conditions, then, by (b.2), it is also $O(n^{1/2})$-bounded. Moreover, they also imply that $\lambda = \sum_{i\geq 1} i\lambda_i$. Janson and Luczak in [18], showed that one can prove a variant of Theorem 7 obtained by replacing the **MR**-conditions by the **JL**-conditions.[5] They also note that if $\lambda_2 = 1$, then the criterion based on $Q(\mathfrak{D})$ does not apply. Our results completely describe the case $\lambda_2 = 1$.

A sequence of degree sequence $\mathfrak{D}$ satisfies the **BR**-*conditions* if

(c.1) it is feasible, smooth and sparse, and

(c.2) $\sum_{i\geq 3} \lambda_i > 0$, and

(c.3) $\lambda = \sum_{i\geq 1} i\lambda_i$.

Bollobás and Riordan in [6] proved a version of Theorem 7 for sequences of degree sequences obtained by replacing the **MR**-conditions by the **BR**-conditions.[6]

[5] Their result gives convergence in probability of the proportion of vertices in the giant component and they also consider the case $Q(\mathfrak{D}) = 0$.

[6] They also proved some results on the distribution of the order of the largest component and also consider the case $Q(\mathfrak{D})) = 0$.

Theorem 7 and its extensions provide easy-to-use criteria for the existence of a giant component and have been widely used by many researchers in the area of complex networks [2, 5, 28]. However, the technical conditions on $\mathfrak{D}$ to which they can be applied, restrict its applicability, seem to be artificial and are only required due to the nature of their proof. As we observed in the introduction, many real-world networks do not satisfy them. For this reason, researchers have developed both ad-hoc approaches for proving results for specific types of degree sequences and variants of the Molloy-Reed result which require different sets of technical conditions to be satisfied.

An early example of an ad-hoc approach is the work of Aiello, Chung and Lu on Power-Law Random Graphs [1]. They introduce a model depending on two parameters $\alpha, \beta > 0$ that define a degree sequence satisfying $n_i = \lfloor e^{\alpha} i^{-\beta} \rfloor$. We should think about these parameters as follows: α is typically large and determines the order of the graph (we always have $\alpha = \Theta(\log n)$), and β is a fixed constant that determines the power-decay of the degree distribution. Among other results, the authors prove that there exists β_0, such that if $\beta > \beta_0$ the probability there is a component of linear order is $o(1)$ and if $\beta < \beta_0$ the probability there is a component of linear order is $1 - o(1)$. Here, the previous conditions are only satisfied for certain values of β and the authors need to do additional work to determine when a giant component exists for other values of β. Theorem 3 also implies their results on the existence of a giant component in the model of Power-Law Random Graphs (see [19]).

1.5 Future Directions

Beginning with the early results of Molloy and Reed, the study of the giant component in random graphs with prescribed degree sequence has attracted a lot of attention. Directions of study include determining the asymptotic order of the largest component in the subcritical regime or estimating the order of the second largest component in both regimes [6, 15, 18, 20, 21, 25, 31]. It would be interesting to extend these known results to arbitrary degree sequences.

For example, Theorem 1 and 2 precisely describe the appearance of a giant component when the degree sequence is well-behaved. While bounds on the constant γ in terms of δ and ϵ respectively, may follow from their respective proofs, these bounds are probably not of the right order of magnitude. Molloy and Reed in [25], precisely determined this dependence for sequences of degree sequences that satisfy the **MR**-conditions. Precise constants are also given in [6, 13, 18]. We wonder whether it is possible to determine the precise dependence on the parameters for arbitrary degree sequences. It is likely that our methods can be used to find this dependence and to determine the order of the second largest component when a giant one exists.

Another direction is the study of site and bond percolation in $G(\mathcal{D})$ for arbitrary degree sequences $\mathcal{D}$. This problem has been already approached for sequences of degree sequences that satisfy certain conditions similar to the ones presented in Section 1.4 [12, 16, 31]. In particular, our approach might be useful to answer a question of Nachmias and Peres on the percolation threshold for random d-regular graphs when $d \to \infty$ as $n \to \infty$ [27].

Motivated by some applications in peer-to-peer networks (see, e.g. [7]), one can study efficient sampling of the random graph $G(\mathcal{D})$. Cooper et al. [8] showed that the switching chain rapidly mixes for d-regular graphs for every $3 \leq d \leq n - 1$. Greenhill [14] recently extended this result to $G(\mathcal{D})$, but, due to some technical reasons, this result only holds if the maximum degree in $\mathcal{D}$ is small enough.

Many other basic properties of $G(\mathcal{D})$, such as determining its diameter [11, 32, 33] or the existence of giant cores [10, 17], have already been studied for certain sequences of degree sequences. We believe that our method can help to extend these results to arbitrary degree sequences.

2 A Proof Sketch

2.1 The Approach

The proofs of Theorem 4, 5 and 6 are simpler than the remaining proofs (see [19]). By applying these theorems, we see that in order to prove Theorem 1 and 2 it is enough to prove the following results:

Theorem 8. *For any function $\delta \to 0$ as $n \to \infty$, for every $\gamma > 0$, if $\mathcal{D}$ is a well-behaved degree sequence with $R_{\mathcal{D}} \leq \delta(n) M_{\mathcal{D}}$, then the probability that $H(\mathcal{D})$ has a component of size at least $\gamma M_{\mathcal{D}}$ is $o(1)$.*

Theorem 9. *For any positive constant ϵ, there is a $\gamma > 0$, such that if $\mathcal{D}$ is a well-behaved degree sequence with $R_{\mathcal{D}} \geq \epsilon M_{\mathcal{D}}$, then the probability that $H(\mathcal{D})$ has a component of size at least $\gamma M_{\mathcal{D}}$ is $1 - o(1)$.*

The proofs of both theorems analyse an exploration process similar to the one discussed in Section 1.1 by combining probabilistic tools with a combinatorial switching argument. However, we will focus on the edges of $H(\mathcal{D})$ rather than the ones of $G(\mathcal{D})$. Again, we will need to bound the expected increase of the number of open edges throughout the process and prove that the (random) increase is highly concentrated around its expected value. In order to do so, we will need to bound the probability that the next vertex of $H(\mathcal{D})$ explored in the process, is a specific vertex w. One of the key applications of our combinatorial switching technique will be to estimate this probability and show that it is approximately proportional to the degree of w, as in Section 1.1.

Crucial to this approach is that the degrees of the vertices explored throughout the process are not too high. Standard arguments for proving concentration of a random variable require that the change at each step is relatively small. This translates precisely to an upper bound on the maximum degree of the explored graph. Furthermore, without such a bound on the maximum degree, we cannot obtain good bounds on the probability that a certain vertex w is the next vertex explored in the process. So, a second key ingredient in our proofs will be a preprocessing step which allows us to handle the vertices of high degree, ensuring that we will not encounter them in our exploration process.

2.2 The Exploration Process

We consider a variant of the exploration process where we start our exploration at a non-empty set S_0 of vertices of $H(\mathcal{D})$, rather than at just one vertex.

Thus, we see that the exploration takes $|V(H(\mathcal{D})) \setminus S_0|$ steps and produces sets

$$S_0 \subset S_1 \subset S_2 \subset \ldots \subset S_{|V(H(\mathcal{D}))\setminus S_0|} ,$$

where $w_t = S_t \setminus S_{t-1}$ is either a neighbour of a vertex v_t of S_{t-1} or is a randomly chosen vertex in $V(H(\mathcal{D})) \setminus S_{t-1}$ if there are no edges between S_{t-1} and $V(H(\mathcal{D})) \setminus S_{t-1}$.

To specify our exploration process precisely, we need to describe how we choose v_t and w_t. To aid in this process, for each vertex $v \in V(H(\mathcal{D}))$ we will choose a uniformly random permutation of its adjacency list in $G(\mathcal{D})$. For this purpose, an *input* of our exploration process consists of a graph G equipped with an ordering of its adjacency lists for all vertices $v \in V(H(\mathcal{D}))$. Applying the method of deferred decisions (cf. Section 2.4 in [26]), we can generate these random linear orders as we go along with our process. We note that this yields, in a natural manner, an ordering of the non-loop edges of $H(\mathcal{D})$ which have the vertex v as an endpoint. If there are no edges between

S_{t-1} and $V(H(\mathcal{D})) \setminus S_{t-1}$, we choose each vertex of $V(H(\mathcal{D})) \setminus S_{t-1}$ to be w_t with probability proportional to its degree. Otherwise we choose the smallest vertex v_t of S_{t-1} (with respect to the natural order in $[n]$), which has a neighbour in $V(H(\mathcal{D})) \setminus S_{t-1}$. We expose the edge of $H(\mathcal{D})$ from v_t to $V(H(\mathcal{D})) \setminus S_{t-1}$ which appears first in our random ordering and let w_t be its other endpoint. Furthermore, we expose all the edges of $H(\mathcal{D})$ from w_t to $S_{t-1} \setminus \{v_t\}$ as well as the loops incident to w_t. Finally, we expose the paths of $G(\mathcal{D})$ corresponding to the edges of $H(\mathcal{D})$ which we have just exposed and the position in the random permutation of the adjacency list of w_t in $G(\mathcal{D})$ of the edges we have just exposed.

Thus, after t iterations of our exploration process we have exposed

- the subgraph of $H(\mathcal{D})$ induced by S_t,
- the paths of $G(\mathcal{D})$ corresponding to the exposed edges of $H(\mathcal{D})$, and
- where each initial and final edge of such a path appears in the random permutation of the adjacency list of its endpoints which are also endpoints of the path.

We refer to this set of information as the *configuration $\mathcal{C}_t$ at time t*. A configuration can also be understood as a set of inputs. During our analysis of the exploration process, we will consider all the probabilities of events conditional on the current configuration.

An important parameter for our exploration process is the number X_t of edges of $H(\mathcal{D})$ between S_t and $V(H(\mathcal{D})) \setminus S_t$. We note that if $X_t = 0$, then S_t is the union of some components of $H(\mathcal{D})$ containing all of S_0. We note that if $|S_0| = 1$, then every X_t is a lower bound on the maximum size of a component of $H(\mathcal{D})$ (not necessarily the one containing the vertex in S_0).

We prove Theorem 8 by showing that under its hypotheses for every vertex v of $H(\mathcal{D})$, there is a set $S_0 = S_0(v)$ containing v such that, given we start our exploration process with S_0, the probability that there is a t with $X_t = 0$ for which the number of edges within S_t is at most $\gamma M_{\mathcal{D}}$, is $1 - o(M_{\mathcal{D}}^{-1})$. Since $H(\mathcal{D})$ has at most $2M_{\mathcal{D}}$ vertices, it follows that the probability that $H(\mathcal{D})$ has a component of size at least $\gamma M_{\mathcal{D}}$ is $o(1)$. The set $S_0 \setminus \{v\}$ is a set of highest degree vertices the sum of whose degrees exceeds $R_{\mathcal{D}}$. By the definition of $j_{\mathcal{D}}$ and $R_{\mathcal{D}}$, this implies that, unless $X_0 = 0$, the expectation of $X_1 - X_0$ is negative. We show that, as the process continues, the expectation of $X_t - X_{t-1}$ becomes even smaller. We can prove that the actual change of X_t is highly concentrated around its expectation and hence complete the proof, because S_0 contains all the high degree vertices and so in the analysis of our exploration process we only have to deal with low degree vertices.

We prove (a slight strengthening of) Theorem 9 for graphs without large degree vertices by showing that under its hypotheses and setting S_0 to be a random vertex v chosen with probability proportional to its degree, with probability $1 - o(1)$, there exists some t such that $X_t \geq \gamma M_{\mathcal{D}}$ (and hence there is a component of $H(\mathcal{D})$ of size at least $\gamma M_{\mathcal{D}}$). Key to doing so is that the expected increase of X_t is a positive fraction of the increase in the sum of the degrees of the vertices in S_t until this sum approaches $R_{\mathcal{D}}$. To handle the high degree vertices, we expose the edges whose endpoints are in components containing a high degree vertex. If this number of edges is at least a constant fraction of $M_{\mathcal{D}}$, then we can show that in fact all the high degree vertices lie in one component, which therefore contains a constant fraction of the edges of $H(\mathcal{D})$. Otherwise, we show that the conditions of Theorem 9 (slightly relaxed) hold in the remainder of the graph, which has no high degree vertices. We then can find the desired component of $H(\mathcal{D})$ in the remainder of the graph, concluding the proof of Theorem 9.

3 Switching

As mentioned above, the key to extending our branching analysis to arbitrary well-behaved degree sequences is a combinatorial switching argument. In this section, we describe the type of switchings we consider and demonstrate the power of the technique.

Let H be a multigraph. We say a multigraph H' is obtained by *switching* from H on a pair of orientations of distinct edges uv and xy, if H' can be obtained from H by deleting uv and xy, and adding the edges ux and vy. Observe that switching ux and vy in H' yields H. Observe further that if H is simple and we want to ensure that H' is simple, then we must insist that $u \neq x$, $v \neq y$ and, unless $u = y$ or $v = x$, the edges ux and vy are not edges of H.

Switching was introduced in the late 19th century by Petersen [30]. Much later, McKay [23] reintroduced the method to count graphs with prescribed degree sequences and, together with Wormald, used it in the study of random regular graphs. We refer the interested reader to the survey of Wormald on random regular graphs for a short introduction to the method [35].

In this paper we will consider standard switchings as well as a particular extension of them. This extension concerns pairs consisting of a simple graph G and the multigraph H_G obtained from G by deleting its cyclic components and suppressing the vertices of degree 2 in the non-cyclic ones. For certain switchings of H_G which yield H', our extension constructs a simple graph G' from G such that $H_{G'} = H'$. We now describe for which switchings in H_G we can obtain such an H' and how we do so.

Our extension considers directed walks (either a path or a cycle) of G which correspond to (oriented) edges in H_G, (note that an edge of H_G corresponds to exactly two such directed walks, even if it is a loop and hence has only one orientation). We can switch on an ordered pair of such directed walks in G, corresponding to an ordered pair of orientated distinct edges $e_1 = uv$ and $e_2 = xy$ of H_G, such that none of the following hold:

(i) there is an edge of G between u and x which forms neither e_1 nor e_2, and the walk corresponding to e_1 has one edge,

(ii) there is an edge of G between v and y which forms neither e_1 nor e_2 and the walk corresponding to e_2 has one edge,

(iii) $u = x$ and the directed walk corresponding to e_1 has at most two edges, or

(iv) $v = y$ and the directed walk corresponding to e_2 has at most two edges.

To do so, let $u = w_0, w_1, \ldots, w_r = v$ be the directed walk corresponding to e_1 and let $x = z_0, z_1, \ldots, z_s = y$ be the directed walk corresponding to e_2. We delete the edges $w_{r-1}v$ and xz_1 and add the edges $w_{r-1}x$ and vz_1.

We note that (i)-(iv) ensure that we obtain a simple graph G'. Furthermore, we have that $H_{G'}$ is obtained from H_G by switching on uv and xy. We remark further that if we reverse both the ordering of the edges and the orientation of both edges, we always obtain the same graph G'; that is, it is equivalent to switch the ordered pair (uv, xy) or the ordered pair (yx, vu). Therefore, given two walks between u and v and between x and y (either paths or cycles) of G, we always consider the four following possible switches: (uv, xy), (uv, yx), (vu, xy) and (vu, yx). We note that some of these choices might give rise to the same graph G'. However, we consider each of them as a valid switch since it will be simpler to count them considering these multiplicities.

Given any two disjoint sets of (multi)graphs $\mathcal{A}$ and $\mathcal{B}$, we can build an auxiliary bipartite graph with vertex set $\mathcal{A} \cup \mathcal{B}$ where we add an edge between $H \in \mathcal{A}$ and $H' \in \mathcal{B}$ for every (extended) switching that transforms H into H'. This definition is symmetric. We can also consider subgraphs

of this auxiliary graph where we only add an edge if the switching satisfies some special property. Given a lower bound $d_{\mathcal{A}}$ on the degrees in $\mathcal{A}$ and an upper bound $d_{\mathcal{B}}$ on the degrees in $\mathcal{B}$, we obtain immediately that $|\mathcal{A}| \leq \frac{d_{\mathcal{B}}}{d_{\mathcal{A}}}|\mathcal{B}|$. We frequently use this fact without explicitly referring to it.

To illustrate our method, we show here that if $M_{\mathcal{D}}$ is large with respect to the number of vertices, then there exists a component containing most of the vertices.

Lemma 10. *If $M_{\mathcal{D}} \geq n \log\log n$, then the probability that $G(\mathcal{D})$ has a component of order $(1-o(1))n$ is $1 - o(1)$.*

In proving the lemma, we will need the following straightforward result on 2-edge cuts of graphs. We defer its proof to the end of the section.

Lemma 11. *The number of pairs of orientations of edges uv, xy in a graph G of order n, such that by switching on uv and xy we obtain a graph with one more component than G, is at most $8n^2$.*

Proof of Lemma 10. We can assume n is large enough to satisfy an inequality stated below since the lemma makes a statement about asymptotic behaviour. Let $K = \lfloor (1 - \frac{1}{\sqrt{\log\log n}})n \rfloor$. For every $k \geq 1$, let $\mathcal{F}_k$ be the event that $G(\mathcal{D})$ has exactly k components and let $\mathcal{F}'_k$ be the event that G is in $\mathcal{F}_k$ and that all components of G have order at most K. Denote by $\mathcal{F}' = \cup_{k\geq 2}\mathcal{F}'_k$. Our goal is to show that $\mathbb{P}[\mathcal{F}'] = o(1)$. If so, with high probability G has a component of order larger than K. Observe that if one proves for some $f(n)$ which is $o(1)$ that $\mathbb{P}[\mathcal{F}'_{k+1}] \leq f(n)\mathbb{P}[\mathcal{F}_k]$, for every $k \geq 1$, then $\mathbb{P}[\mathcal{F}'] = \sum_{k\geq 1} \mathbb{P}[\mathcal{F}'_{k+1}] \leq f(n)\left(\sum_{k\geq 1} \mathbb{P}[\mathcal{F}_k]\right) = f(n) = o(1)$. We adopt this approach with $f(n) = \frac{16}{\sqrt{\log\log n}}$.

Fix $k \geq 1$. Now suppose that there exist s^+ and s^- such that for every G in $\mathcal{F}_k$, there are at most s^+ switchings that transform G into a graph in $\mathcal{F}'_{k+1}$, and for every graph G in $\mathcal{F}'_{k+1}$, there are at least s^- switchings that transform G into a graph in $\mathcal{F}_k$. Then,

$$\mathbb{P}[\mathcal{F}'_{k+1}]s^- \leq \mathbb{P}[\mathcal{F}_k]s^+ \ .$$

Let us now obtain some values for s^+ and s^-. On the one hand, applying Lemma 11, we can choose

$$s^+ = 8n^2 \ .$$

On the other hand, if G is in $\mathcal{F}'_{k+1}$, in order to merge two components it is enough to perform a switching between an oriented non-cut edge (at least $M_{\mathcal{D}} - 2n \geq (\log\log n - 2)n$ choices) and any other oriented edge not in the same component as the first one (since G has minimum degree at least 1 and the largest component has order at most K, there are at least $n - K$ choices). Since n is large, we can choose

$$s^- = (\log\log n - 2)n \cdot (n - K) \geq \frac{\sqrt{\log\log n}}{2} n^2 \ .$$

From the previous bounds, we obtain the desired result

$$\mathbb{P}[\mathcal{F}'_{k+1}] \leq \frac{s^+}{s^-} \cdot \mathbb{P}[\mathcal{F}_k] \leq \frac{16}{\sqrt{\log\log n}} \cdot \mathbb{P}[\mathcal{F}_k] \ .$$

□

References

[1] W. Aiello, F. Chung, and L. Lu, *A random graph model for massive graphs*, Proceedings of the thirty-second annual ACM Symposium on Theory of Computing (STOC), ACM, 2000, pp. 171–180.

[2] R. Albert and A.-L. Barabási, *Statistical mechanics of complex networks*, Reviews of modern physics **74** (2002), 47.

[3] R. Arratia, A. D. Barbour, and S. Tavaré, *Logarithmic combinatorial structures: a probabilistic approach*, European Mathematical Society, 2003.

[4] A.-L. Barabási and R. Albert, *Emergence of scaling in random networks*, science **286** (1999), no. 5439, 509–512.

[5] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, *Complex networks: Structure and dynamics*, Physics reports **424** (2006), 175–308.

[6] B. Bollobás and O. Riordan, *An old approach to the giant component problem*, J. Combin. Theory (Series B) **113** (2015), 236–260.

[7] V. Bourassa and F. Holt, *SWAN: Small-world wide area networks*, Proceeding of International Conference on Advances in Infrastructures (SSGRR 2003w), L'Aquila, Italy, 2003.

[8] C. Cooper, M. Dyer, and C. Greenhill, *Sampling regular graphs and a peer-to-peer network*, Comb., Probab. Comput. **16** (2007), 557–593.

[9] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *On power-law relationships of the internet topology*, ACM SIGCOMM computer communication review, vol. 29, ACM, 1999, pp. 251–262.

[10] D. Fernholz and V. Ramachandran, *Cores and connectivity in sparse random graphs*, The University of Texas at Austin, Department of Computer Sciences, technical report TR-04-13 (2004).

[11] ———, *The diameter of sparse random graphs*, Random Structures Algorithms **31** (2007), 482–516.

[12] N. Fountoulakis, *Percolation on sparse random graphs with given degree sequence*, Internet Mathematics **4** (2007), 329–356.

[13] N. Fountoulakis and B. Reed, *A general critical condition for the emergence of a giant component in random graphs with given degrees*, Electronic Notes in Discrete Mathematics **34** (2009), 639–645.

[14] C. Greenhill, *The switch markov chain for sampling irregular graphs*, Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2015, pp. 1564–1572.

[15] H. Hatami and M. Molloy, *The scaling window for a random graph with a given degree sequence*, Random Structures Algorithms **41** (2012), 99–123.

[16] S. Janson, *On percolation in random graphs with given vertex degrees*, Electron. J. Probab. **14** (2009), 87–118.

[17] S. Janson and M.J. Luczak, *A simple solution to the k-core problem*, Random Structures Algorithms **30** (2007), 50–62.

[18] _____, *A new approach to the giant component problem*, Random Structures Algorithms **34** (2009), 197–216.

[19] F. Joos, G. Perarnau, D. Rautenbach, and B. Reed, *How to determine if a random graph with a fixed degree sequence has a giant component*, arXiv:1601.03714 (2016).

[20] A. Joseph, *The component sizes of a critical random graph with pre-described degree sequence*, The Annals of Applied Probability **24** (2014), 2560–2594.

[21] M. Kang and T. G. Seierstad, *The critical phase for random graphs with a given degree sequence*, Comb., Probab. Comput. **17** (2008), 67–86.

[22] T. Luczak, *Sparse random graphs with a given degree sequence*, Proceedings of the Symposium on Random Graphs, Poznan, 1989, pp. 165–182.

[23] B. D. McKay, *Subgraphs of random graphs with specified degrees*, vol. 33, 1981, pp. 213–223.

[24] M. Molloy and B. Reed, *A critical point for random graphs with a given degree sequence*, Random Structures Algorithms **6** (1995), 161–180.

[25] _____, *The size of the giant component of a random graph with a given degree sequence*, Comb., Probab. Comput. **7** (1998), 295–305.

[26] _____, *Graph colouring and the probabilistic method*, vol. 23, Springer Science & Business Media, 2013.

[27] A. Nachmias and Y. Peres, *Critical percolation on random regular graphs*, Random Structures Algorithms **36** (2010), 111–148.

[28] M. E. J. Newman, *The structure and function of complex networks*, SIAM review **45** (2003), 167–256.

[29] M. E. J. Newman, D. J. Watts, and S. H. Strogatz, *Random graph models of social networks*, Proceedings of the National Academy of Sciences **99** (2002), no. suppl 1, 2566–2572.

[30] J. Petersen, *Die Theorie der regulären graphs*, Acta Math. **15** (1891), 193–220.

[31] O. Riordan, *The phase transition in the configuration model*, Comb., Probab. Comput. **21** (2012), 265–299.

[32] H. van den Esker, R. van der Hofstad, G. Hooghiemstra, and D. Znamenski, *Distances in random graphs with infinite mean degrees*, Extremes **8** (2005), no. 3, 111–141.

[33] R. van der Hofstad, G. Hooghiemstra, and P. van Mieghem, *Distances in random graphs with finite variance degrees*, Random Structures Algorithms **27** (2005), 76–123.

[34] N. C. Wormald, *Some problems in the enumeration of labelled graphs*, Doctoral Thesis (1980).

[35] _____, *Models of random regular graphs*, Surveys in combinatorics, 1999 (Canterbury), London Math. Soc. Lecture Note Ser., vol. 267, Cambridge Univ. Press, Cambridge, 1999, pp. 239–298.